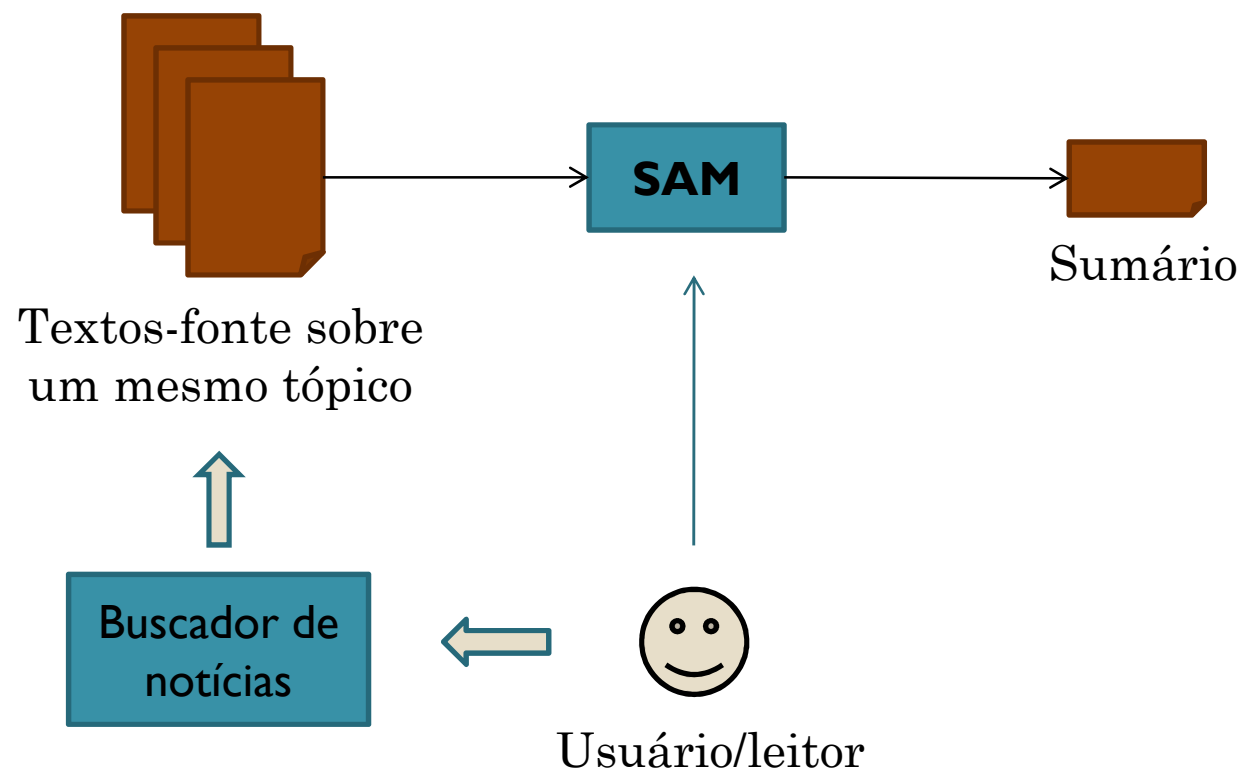


**Interface Web** para o projeto: Sumarização  
Automática Multidocumento para o Português do  
Brasil com Base na Teoria de Estruturação  
Multidocumento CST (*Cross-document Structure  
Theory*)

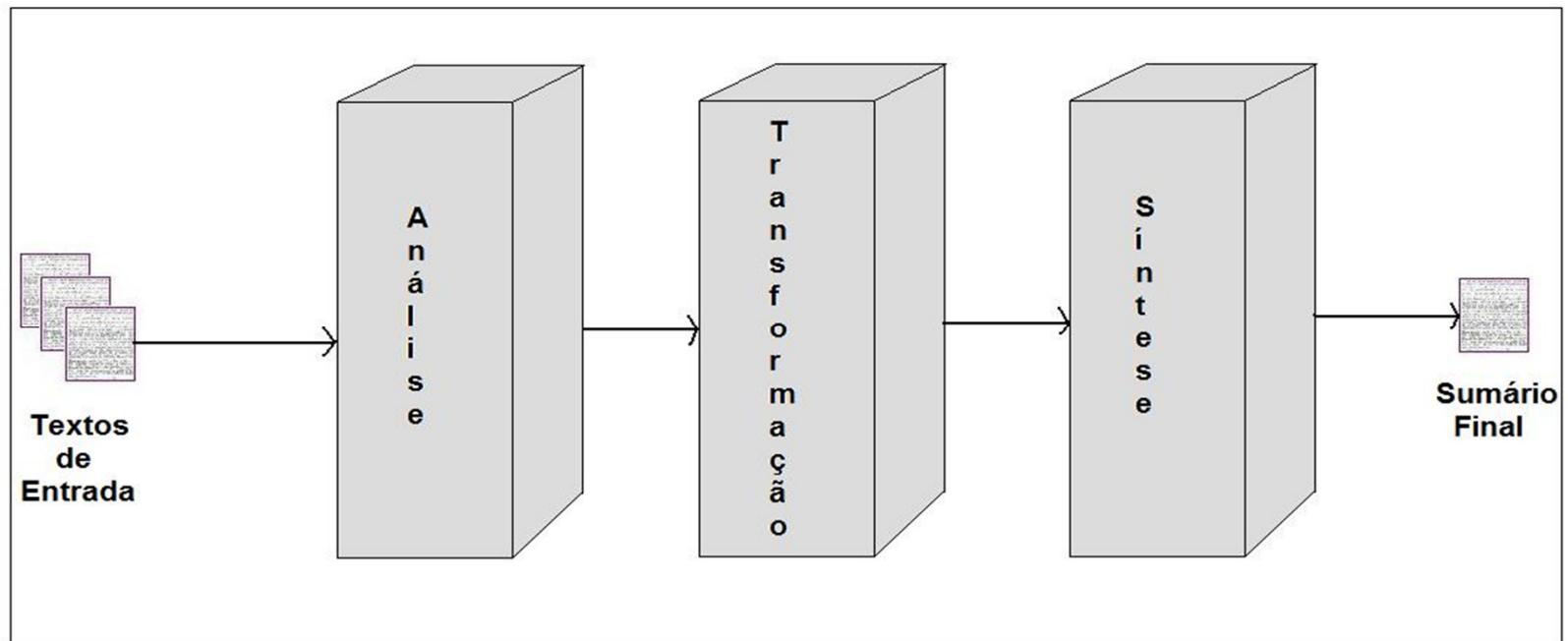
Pedro Paulo Balage Filho

# Introdução

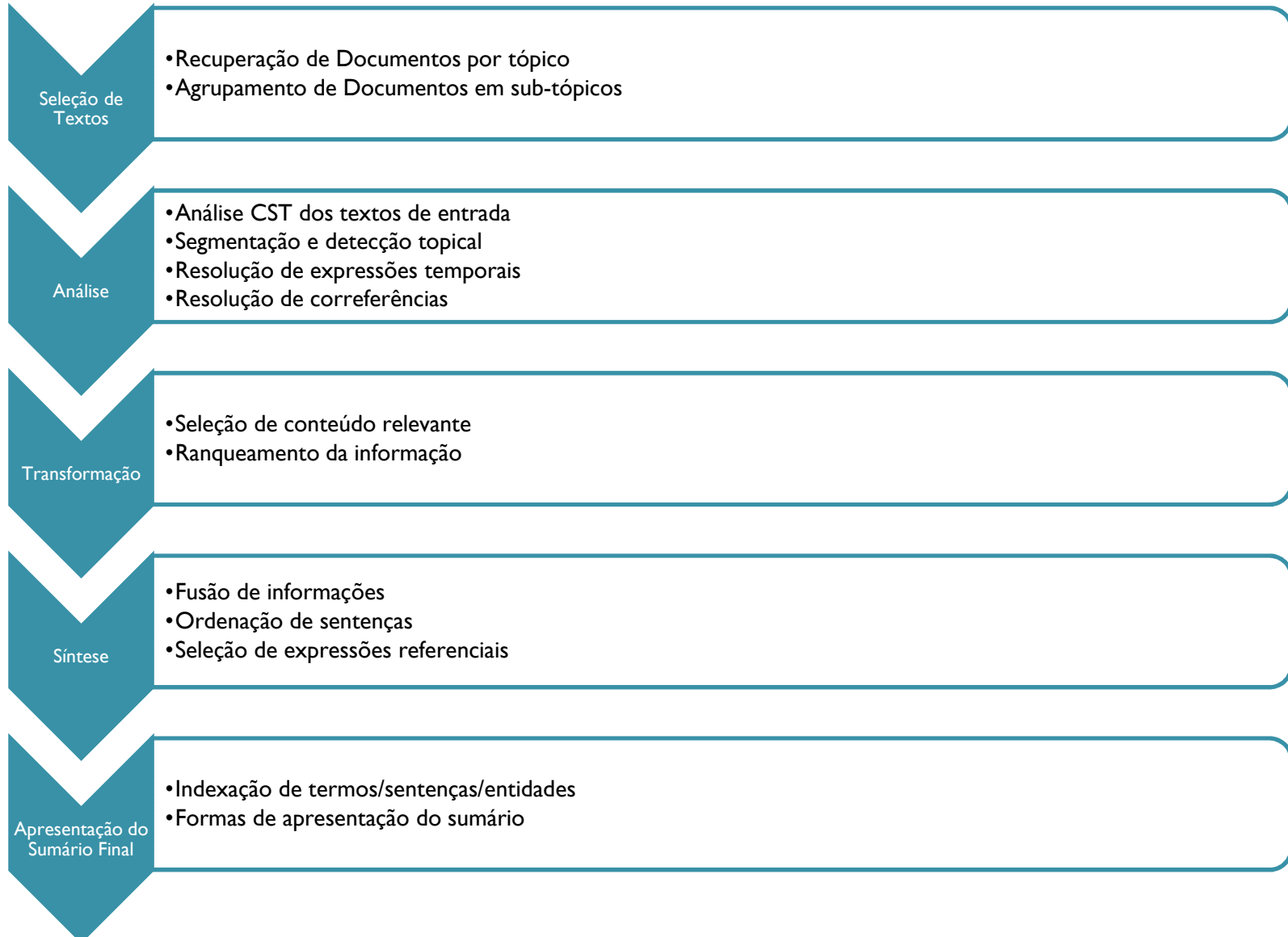
- Cenário do projeto



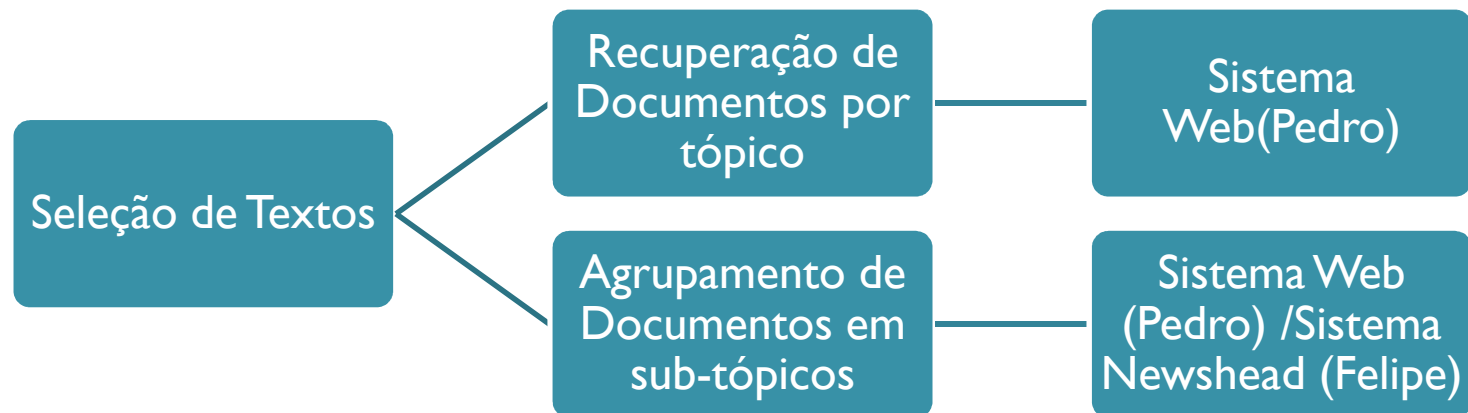
# Arquitetura Genérica de SA



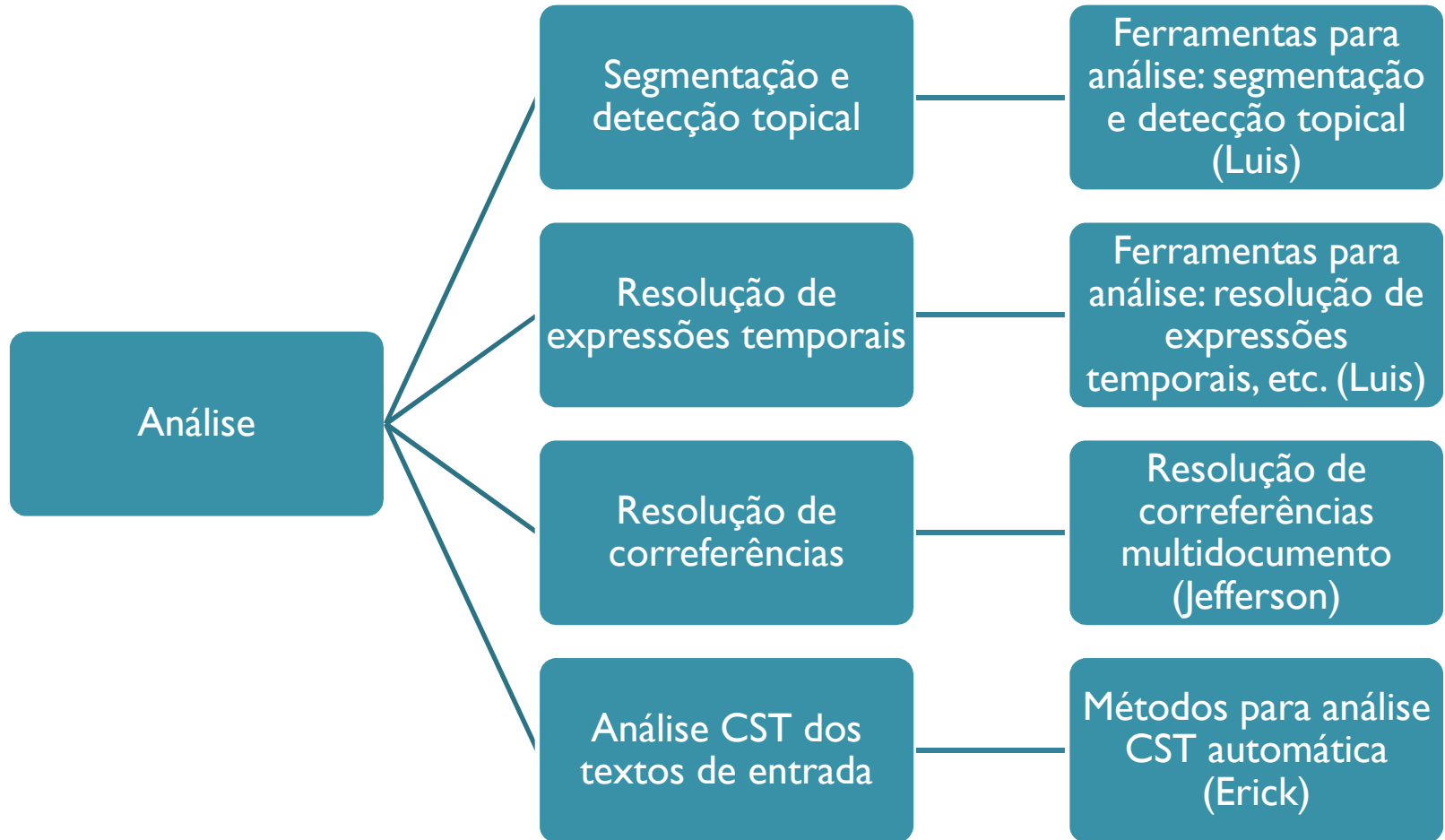
# Arquitetura do Processo de Sumarização



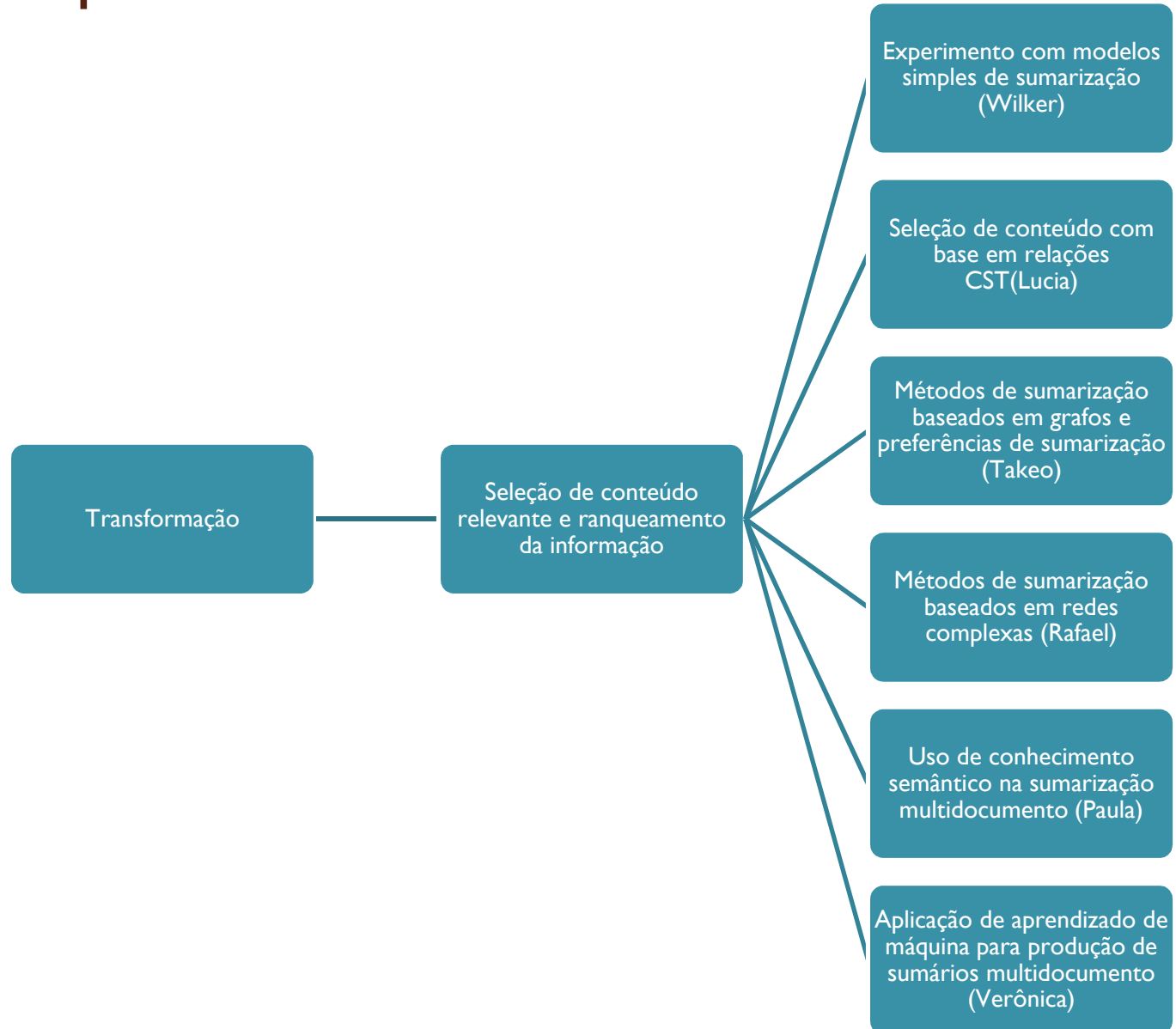
# Arquitetura do Processo de Sumarização



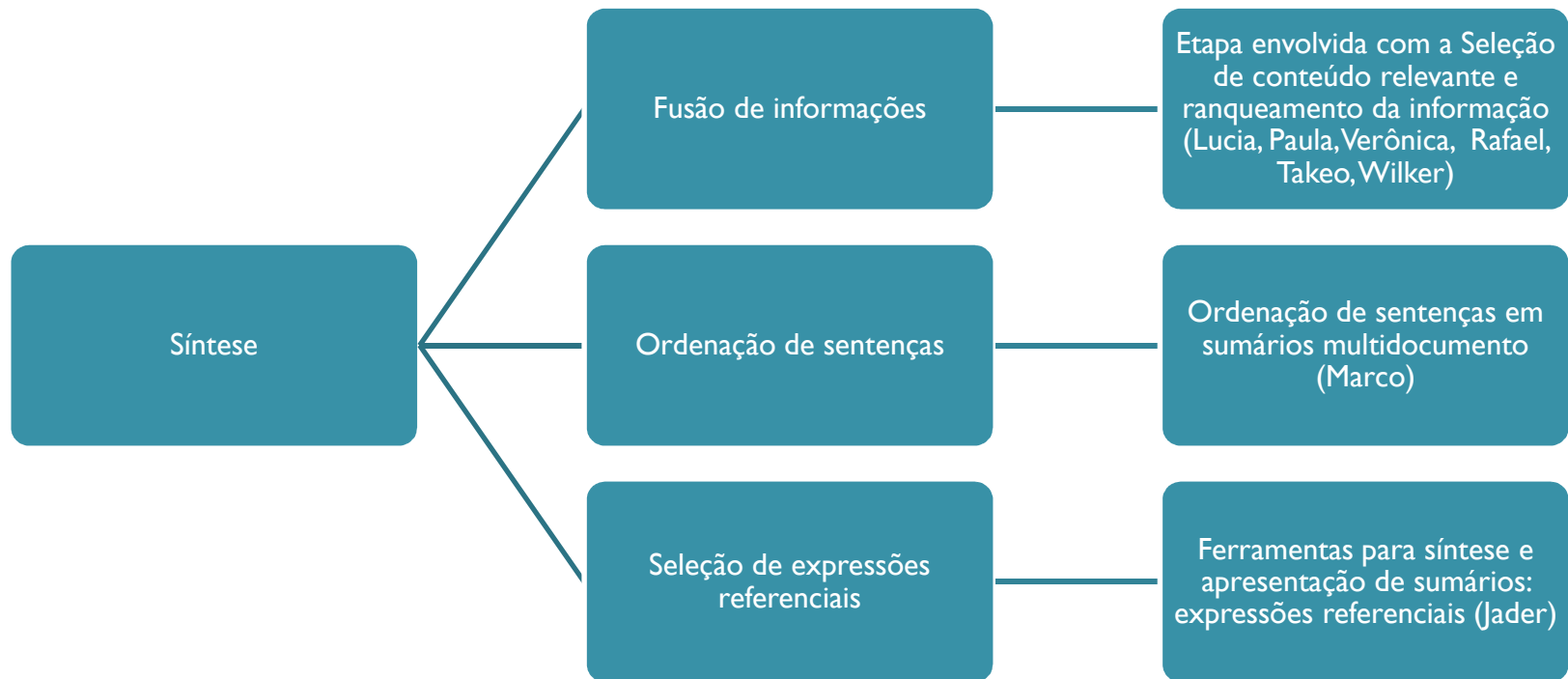
# Arquitetura do Processo de Sumarização



# Arquitetura do Processo de Sumarização



# Arquitetura do Processo de Sumarização



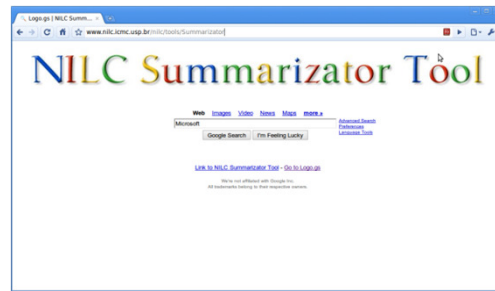


# Arquitetura do Processo de Sumarização

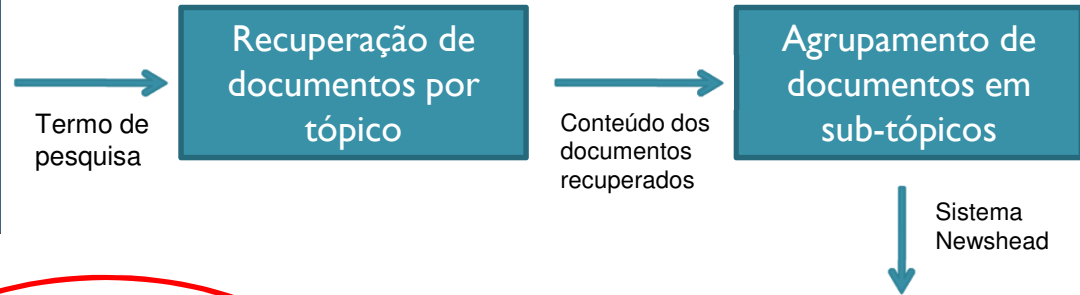


# Arquitetura do Sistema Web

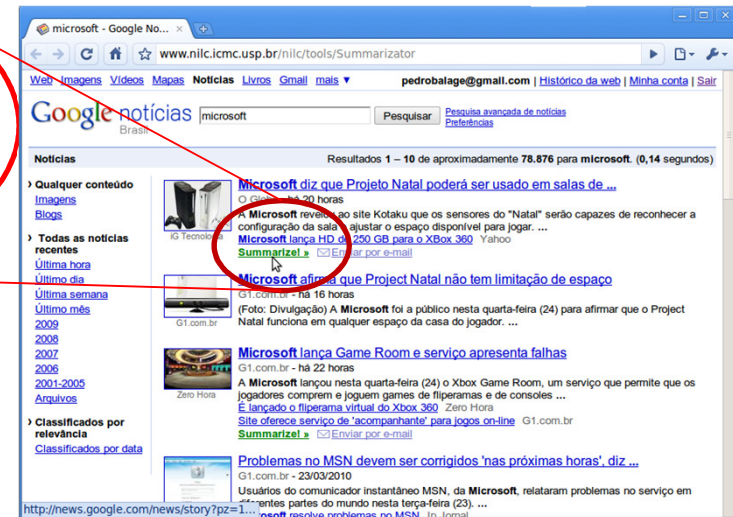
## Seleção de Textos



Tela Principal



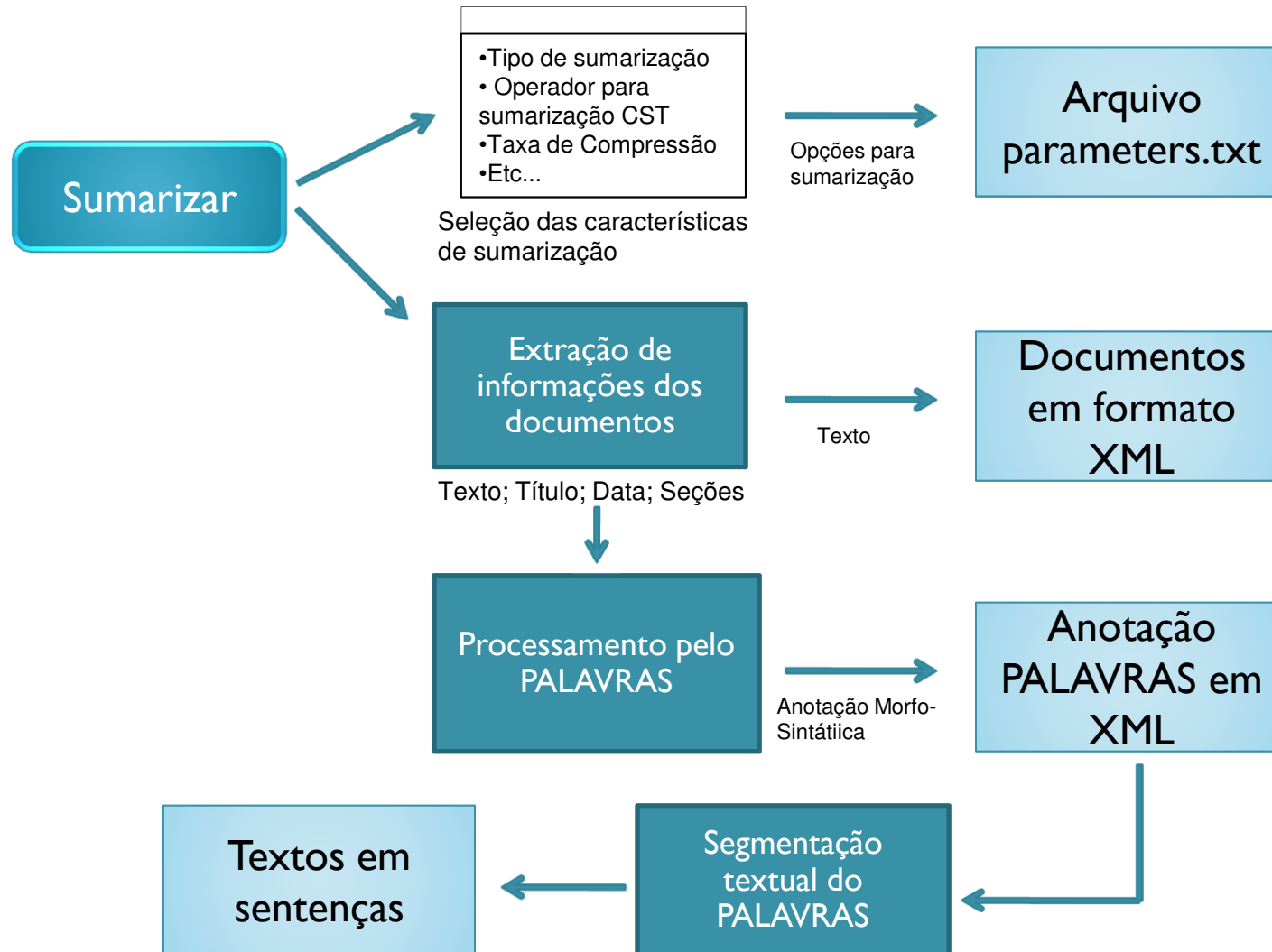
Iteração do usuário com os textos recuperados



Apresentação dos resultados ao usuário

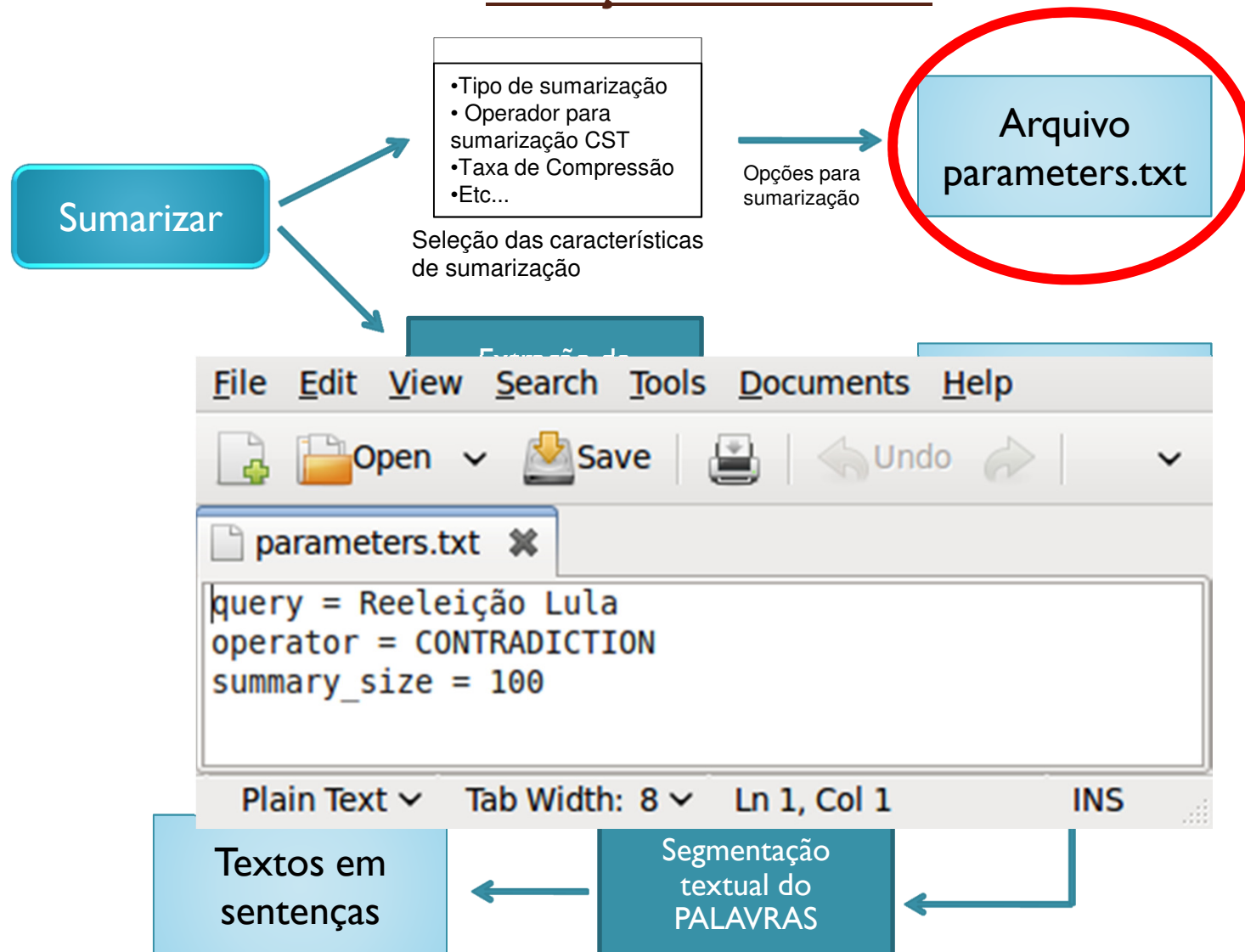
# Arquitetura do Sistema Web

## Seleção de Textos



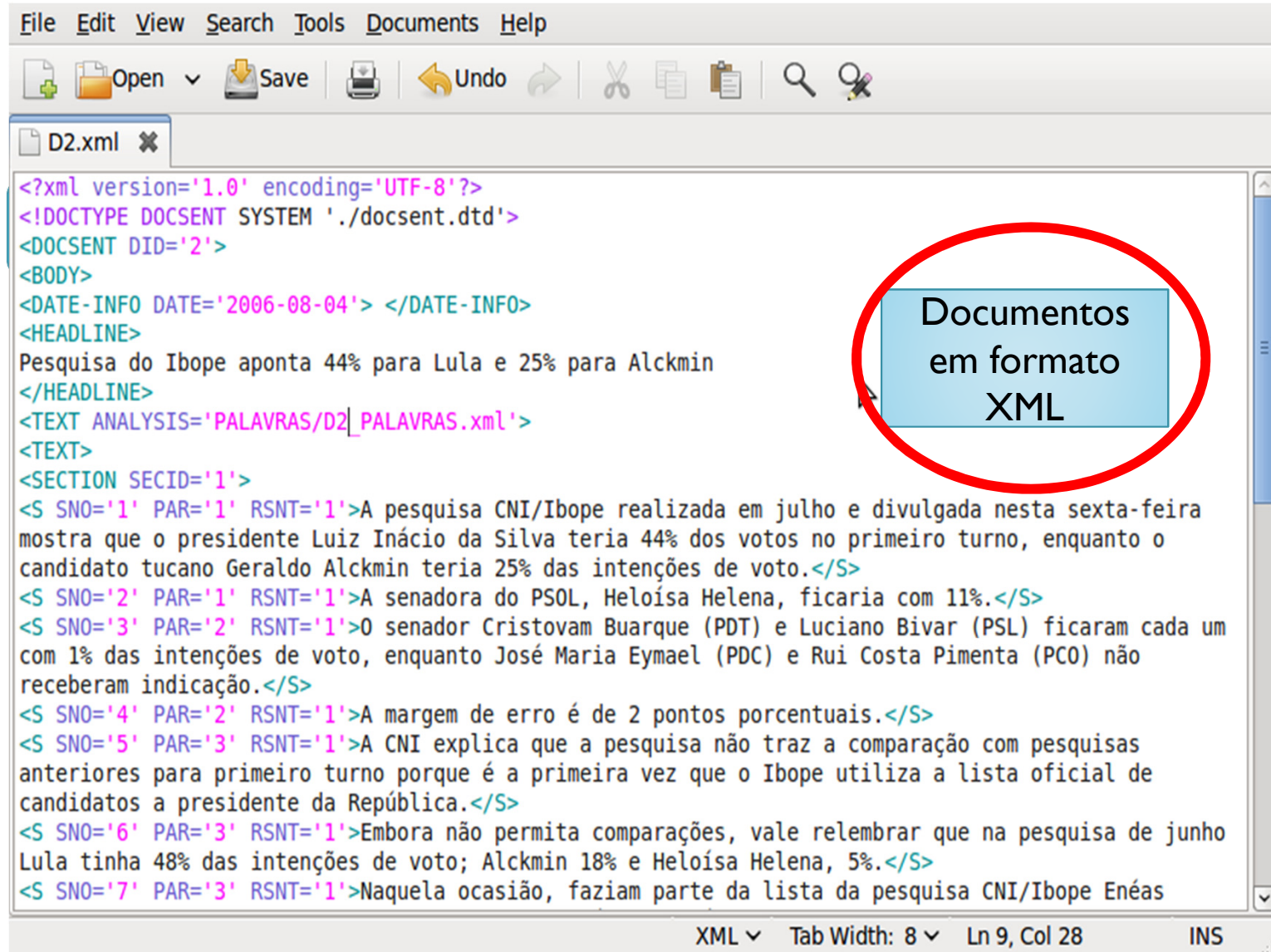
# Arquitetura do Sistema Web

## Seleção de Textos



# Arquitetura do Sistema Web

## Seleção de Textos



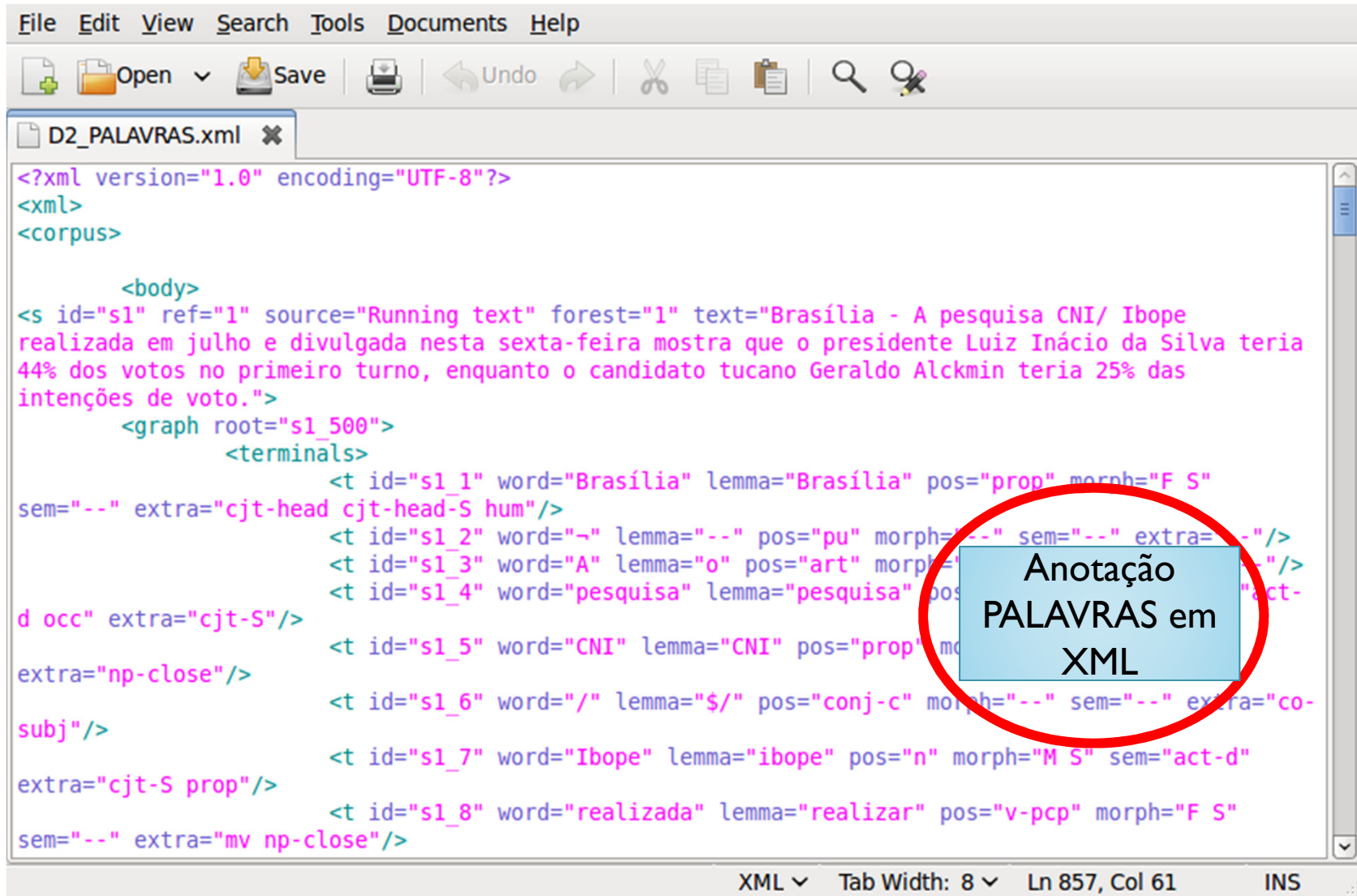
```
File Edit View Search Tools Documents Help
Open Save Undo
D2.xml x
<?xml version='1.0' encoding='UTF-8'?>
<!DOCTYPE DOCSSENT SYSTEM './docsent.dtd'>
<DOCSSENT DID='2'>
<BODY>
<DATE-INFO DATE='2006-08-04'> </DATE-INFO>
<HEADLINE>
Pesquisa do Ibope aponta 44% para Lula e 25% para Alckmin
</HEADLINE>
<TEXT ANALYSIS='PALAVRAS/D2_PALAVRAS.xml'>
<TEXT>
<SECTION SECID='1'>
<S SNO='1' PAR='1' RSNT='1'>A pesquisa CNI/Ibope realizada em julho e divulgada nesta sexta-feira
mostra que o presidente Luiz Inácio da Silva teria 44% dos votos no primeiro turno, enquanto o
candidato tucano Geraldo Alckmin teria 25% das intenções de voto.</S>
<S SNO='2' PAR='1' RSNT='1'>A senadora do PSOL, Heloísa Helena, ficaria com 11%.</S>
<S SNO='3' PAR='2' RSNT='1'>O senador Cristovam Buarque (PDT) e Luciano Bivar (PSL) ficaram cada um
com 1% das intenções de voto, enquanto José Maria Eymael (PDC) e Rui Costa Pimenta (PCO) não
receberam indicação.</S>
<S SNO='4' PAR='2' RSNT='1'>A margem de erro é de 2 pontos porcentuais.</S>
<S SNO='5' PAR='3' RSNT='1'>A CNI explica que a pesquisa não traz a comparação com pesquisas
anteriores para primeiro turno porque é a primeira vez que o Ibope utiliza a lista oficial de
candidatos a presidente da República.</S>
<S SNO='6' PAR='3' RSNT='1'>Embora não permita comparações, vale relembrar que na pesquisa de junho
Lula tinha 48% das intenções de voto; Alckmin 18% e Heloísa Helena, 5%.</S>
<S SNO='7' PAR='3' RSNT='1'>Naquela ocasião, faziam parte da lista da pesquisa CNI/Ibope Enéas
```

Documentos em formato XML

XML Tab Width: 8 Ln 9, Col 28 INS

# Arquitetura do Sistema Web

## Seleção de Textos

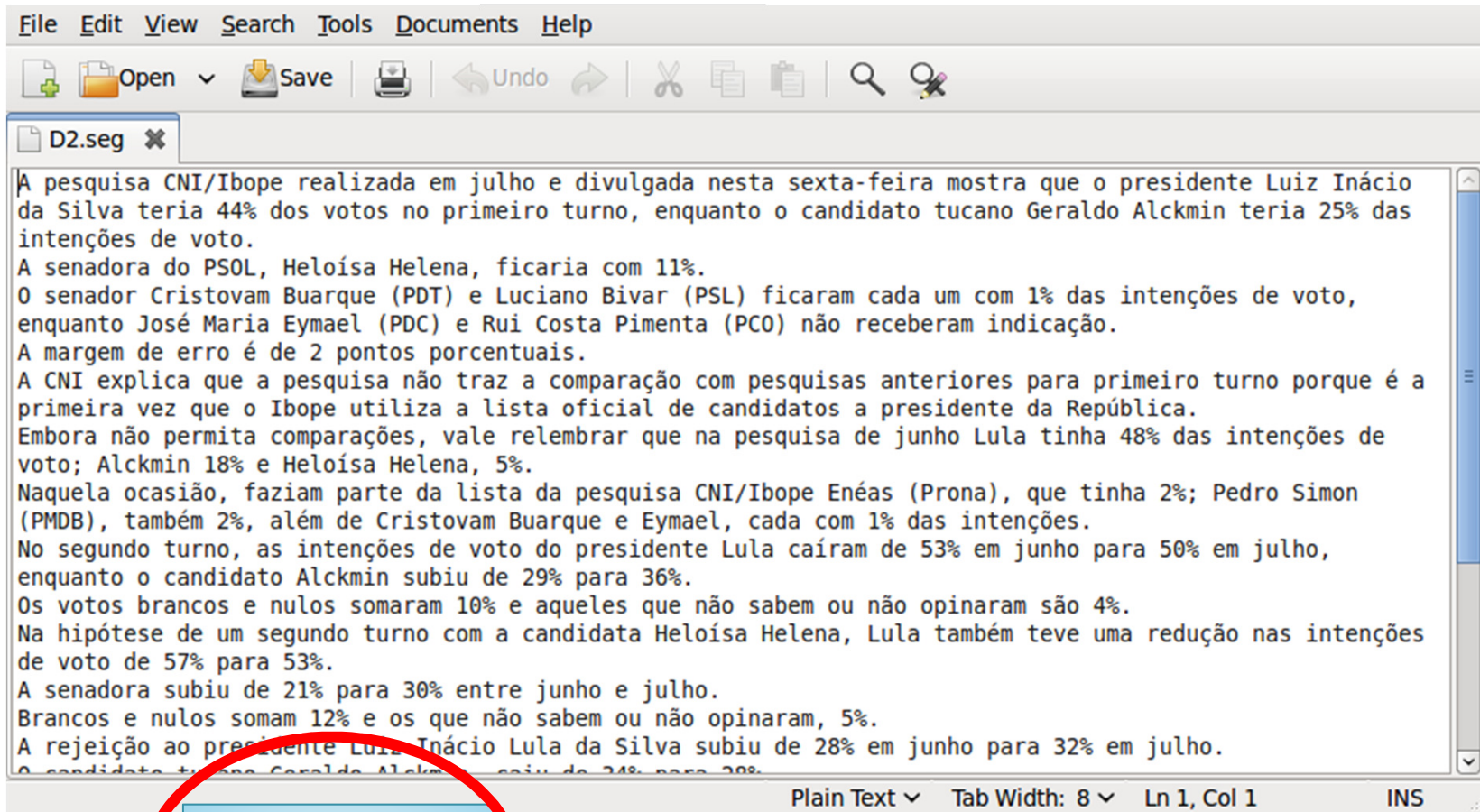


```
File Edit View Search Tools Documents Help
Open Save Undo
D2_PALAVRAS.xml
<?xml version="1.0" encoding="UTF-8"?>
<xml>
<corpus>
  <body>
<s id="s1" ref="1" source="Running text" forest="1" text="Brasília - A pesquisa CNI/ Ibope
realizada em julho e divulgada nesta sexta-feira mostra que o presidente Luiz Inácio da Silva teria
44% dos votos no primeiro turno, enquanto o candidato tucano Geraldo Alckmin teria 25% das
intenções de voto.">
  <graph root="s1_500">
    <terminals>
      <t id="s1_1" word="Brasília" lemma="Brasília" pos="prop" morph="F S"
sem="--" extra="cjt-head cjt-head-S hum"/>
      <t id="s1_2" word="-" lemma="--" pos="pu" morph="--" sem="--" extra="--"/>
      <t id="s1_3" word="A" lemma="o" pos="art" morph="--" sem="--" extra="--"/>
      <t id="s1_4" word="pesquisa" lemma="pesquisa" pos="n" morph="F S" sem="act-
d occ" extra="cjt-S"/>
      <t id="s1_5" word="CNI" lemma="CNI" pos="prop" morph="F S" sem="--" extra="np-close"/>
      <t id="s1_6" word="/" lemma="$/" pos="conj-c" morph="--" sem="--" extra="co-
subj"/>
      <t id="s1_7" word="Ibope" lemma="ibope" pos="n" morph="M S" sem="act-d"
extra="cjt-S prop"/>
      <t id="s1_8" word="realizada" lemma="realizar" pos="v-pcp" morph="F S"
sem="--" extra="mv np-close"/>
    </terminals>
  </graph>
</s>
</body>
</corpus>
</xml>
XML Tab Width: 8 Ln 857, Col 61 INS
```

Anotação  
PALAVRAS em  
XML

# Arquitetura do Sistema Web

## Seleção de Textos

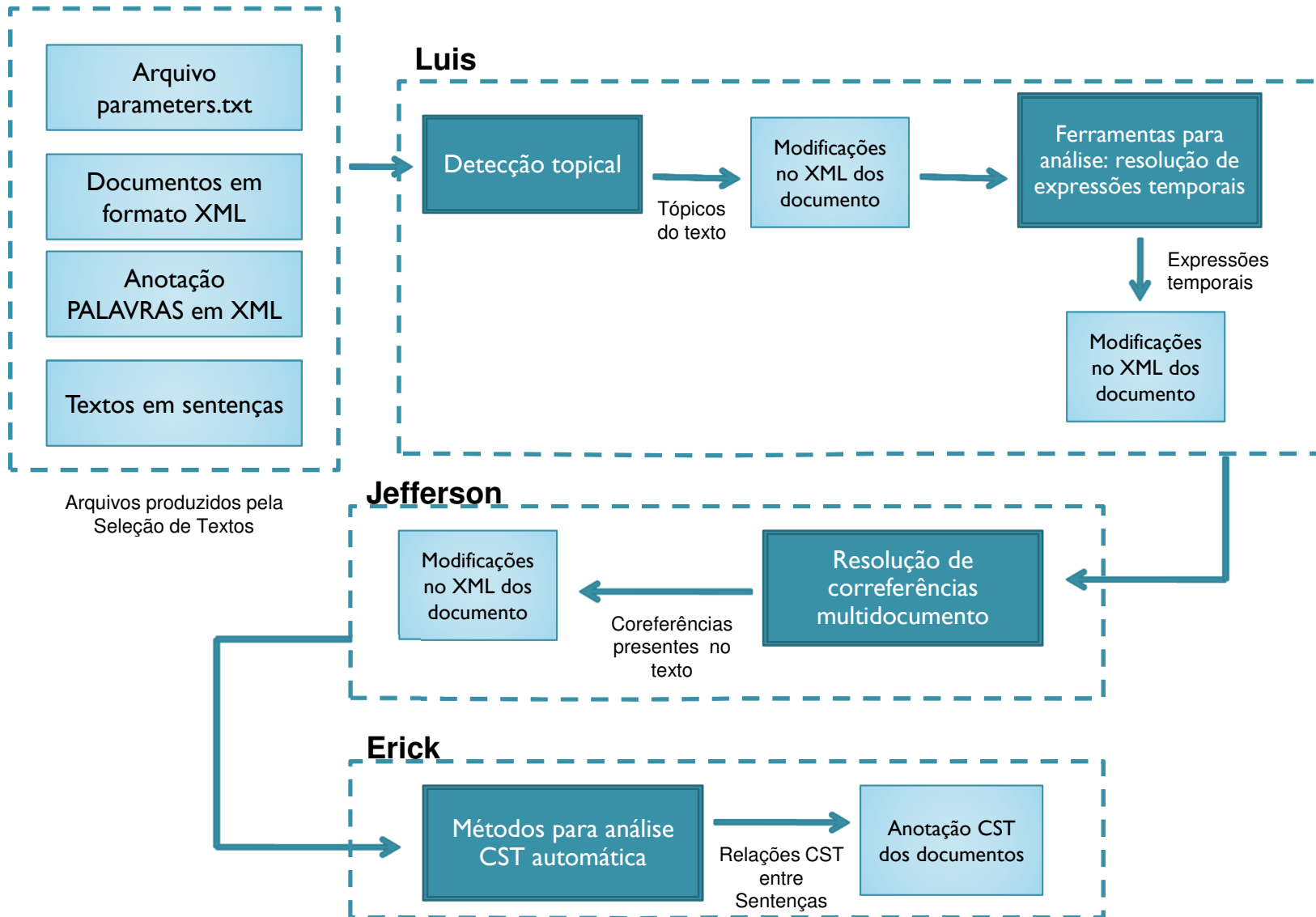


Textos em  
sentenças

Segmentação  
textual do  
PALAVRAS

# Arquitetura do Sistema Web

## Análise





# Arquitetura do Sistema Web

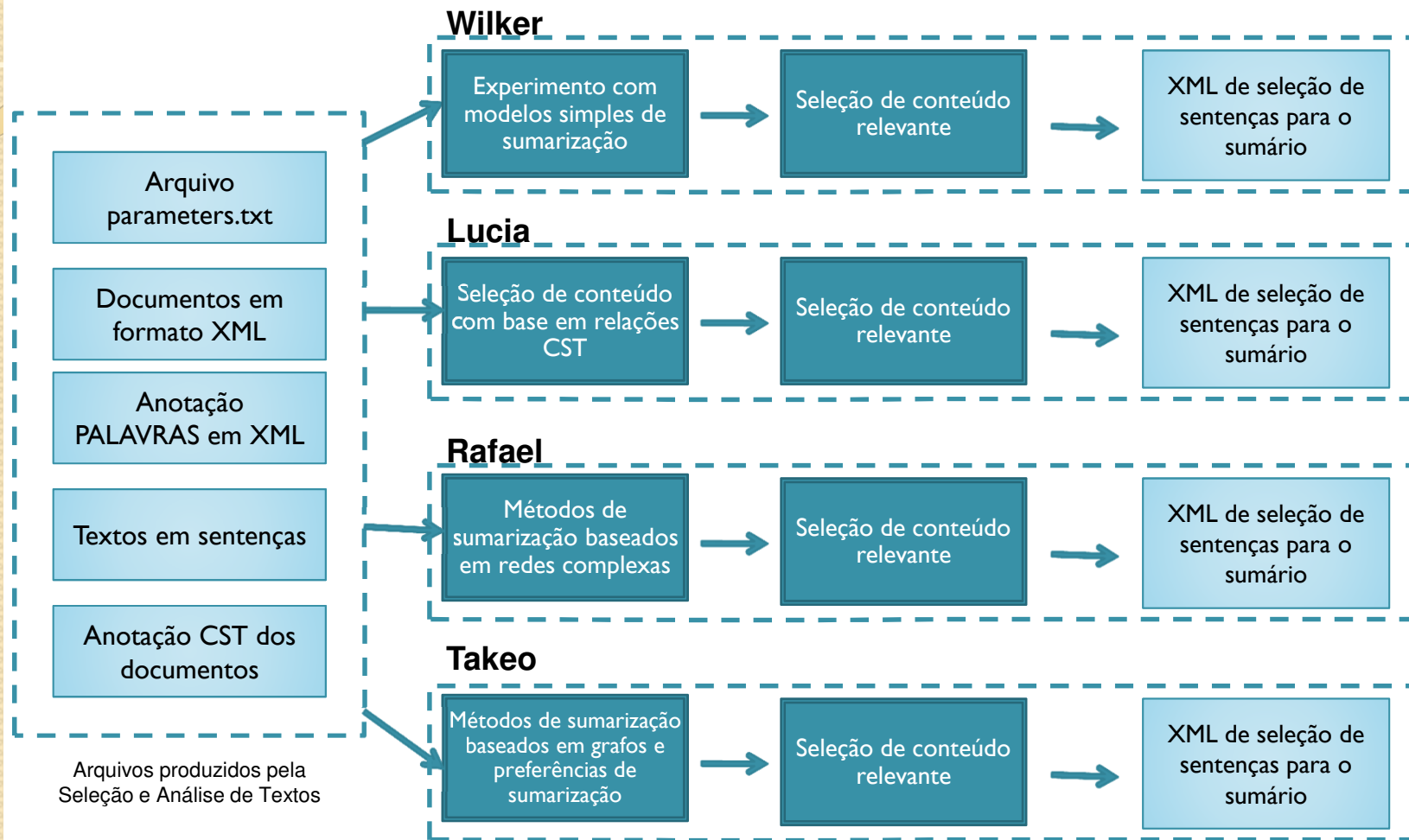
## Análise

```
File Edit View Search Tools Documents Help
Open Save Undo
file.cst
<?xml version='1.0' encoding='UTF-8'?>
<!DOCTYPE TABLE SYSTEM "./sentrel.dtd">
<TABLE>
<R SDID="D2_C2_Estadao.txt.seg" SSENT="1" TDID="D1_C2_Folha.txt.seg" TSENT="1">
<RELATION TYPE="Subsumption" JUDGE="5"/>
</R>
<R SDID="D2_C2_Estadao.txt.seg" SSENT="8" TDID="D1_C2_Folha.txt.seg" TSENT="1">
<RELATION TYPE="Follow-up" JUDGE="5"/>
</R>
<R SDID="D2_C2_Estadao.txt.seg" SSENT="18" TDID="D1_C2_Folha.txt.seg" TSENT="1">
<RELATION TYPE="Elaboration" JUDGE="5"/>
</R>
<R SDID="D1_C2_Folha.txt.seg" SSENT="2" TDID="D2_C2_Estadao.txt.seg" TSENT="2">
<RELATION TYPE="Subsumption" JUDGE="5"/>
</R>
<R SDID="D1_C2_Folha.txt.seg" SSENT="2" TDID="D2_C2_Estadao.txt.seg" TSENT="3">
<RELATION TYPE="Overlap" JUDGE="5"/>
</R>
<R SDID="D2_C2_Estadao.txt.seg" SSENT="10" TDID="D1_C2_Folha.txt.seg" TSENT="2">
<RELATION TYPE="Follow-up" JUDGE="5"/>
</R>
<R SDID="D2_C2_Estadao.txt.seg" SSENT="16" TDID="D1_C2_Folha.txt.seg" TSENT="2">
<RELATION TYPE="Follow-up" JUDGE="5"/>
</R>
```



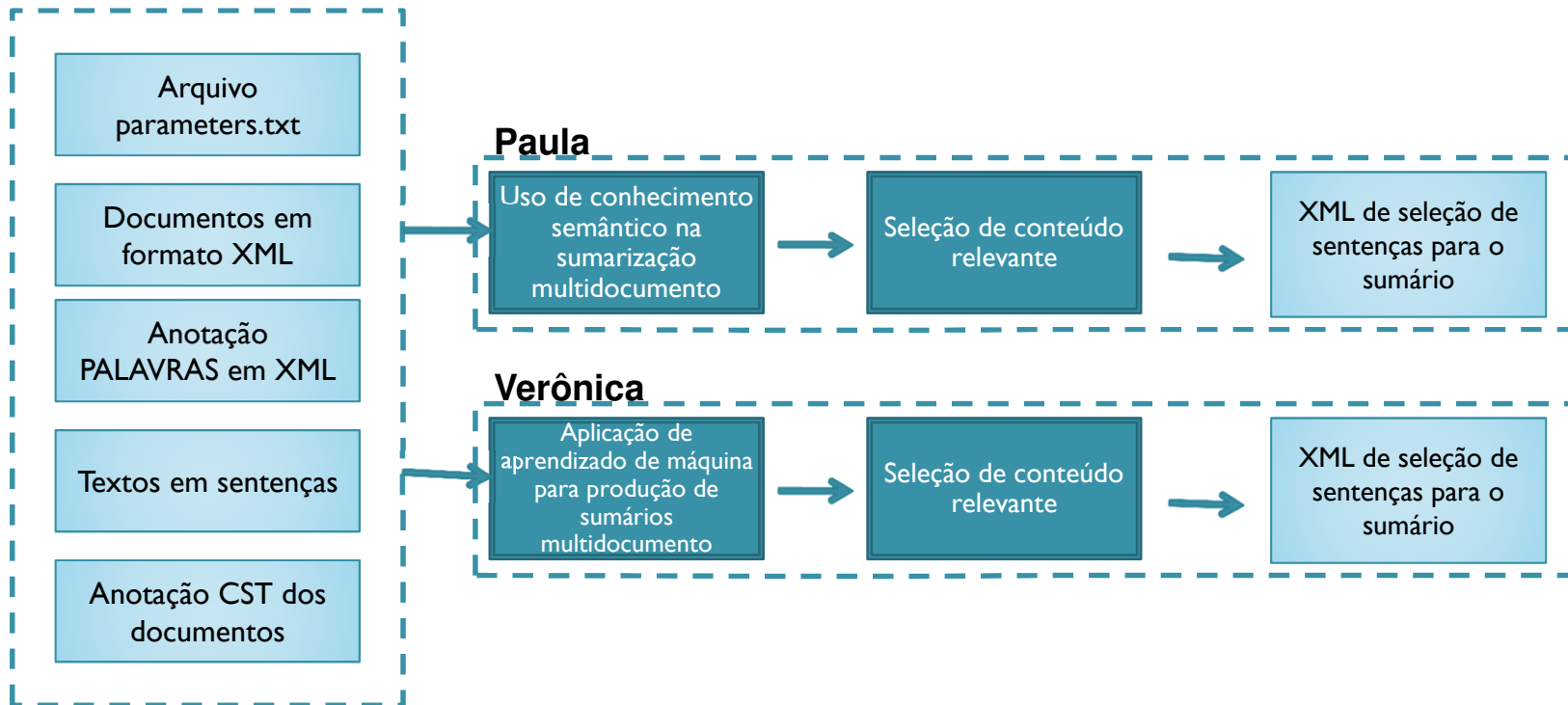
# Arquitetura do Sistema Web

## Transformação e Síntese



# Arquitetura do Sistema Web

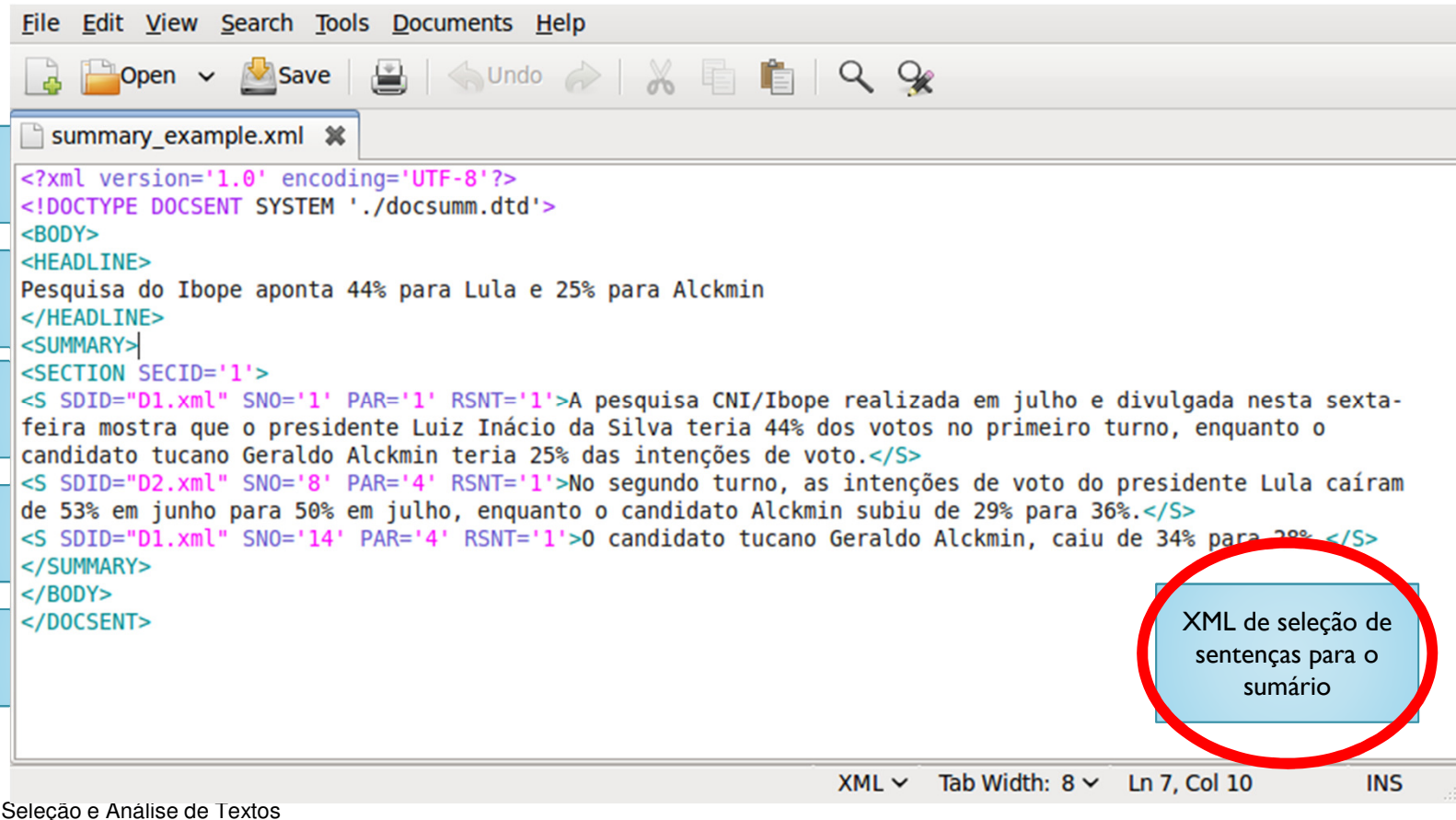
## Transformação e Síntese



Arquivos produzidos pela  
Seleção e Análise de Textos

# Arquitetura do Sistema Web

## Transformação e Síntese



```
File Edit View Search Tools Documents Help
Open Save Undo
summary_example.xml
<?xml version='1.0' encoding='UTF-8'?>
<!DOCTYPE DOCSSENT SYSTEM './docsumm.dtd'>
<BODY>
<HEADLINE>
Pesquisa do Ibope aponta 44% para Lula e 25% para Alckmin
</HEADLINE>
<SUMMARY>
<SECTION SECID='1'>
<S SDID='D1.xml' SNO='1' PAR='1' RSNT='1'>A pesquisa CNI/Ibope realizada em julho e divulgada nesta sexta-
feira mostra que o presidente Luiz Inácio da Silva teria 44% dos votos no primeiro turno, enquanto o
candidato tucano Geraldo Alckmin teria 25% das intenções de voto.</S>
<S SDID='D2.xml' SNO='8' PAR='4' RSNT='1'>No segundo turno, as intenções de voto do presidente Lula caíram
de 53% em junho para 50% em julho, enquanto o candidato Alckmin subiu de 29% para 36%.</S>
<S SDID='D1.xml' SNO='14' PAR='4' RSNT='1'>O candidato tucano Geraldo Alckmin, caiu de 34% para 28% </S>
</SUMMARY>
</BODY>
</DOCSSENT>
```

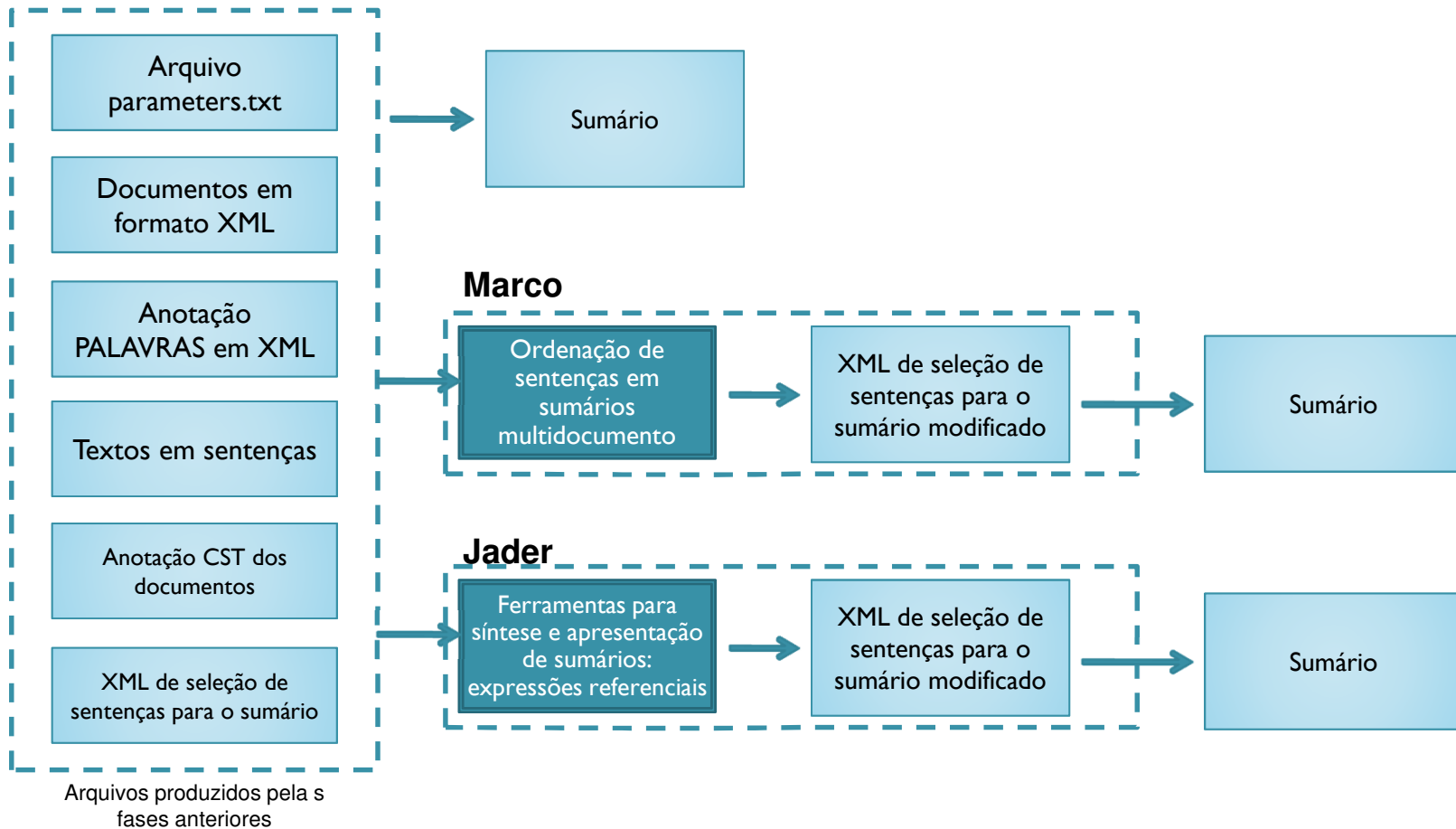
XML de seleção de sentenças para o sumário

XML Tab Width: 8 Ln 7, Col 10 INS

Seleção e Análise de Textos

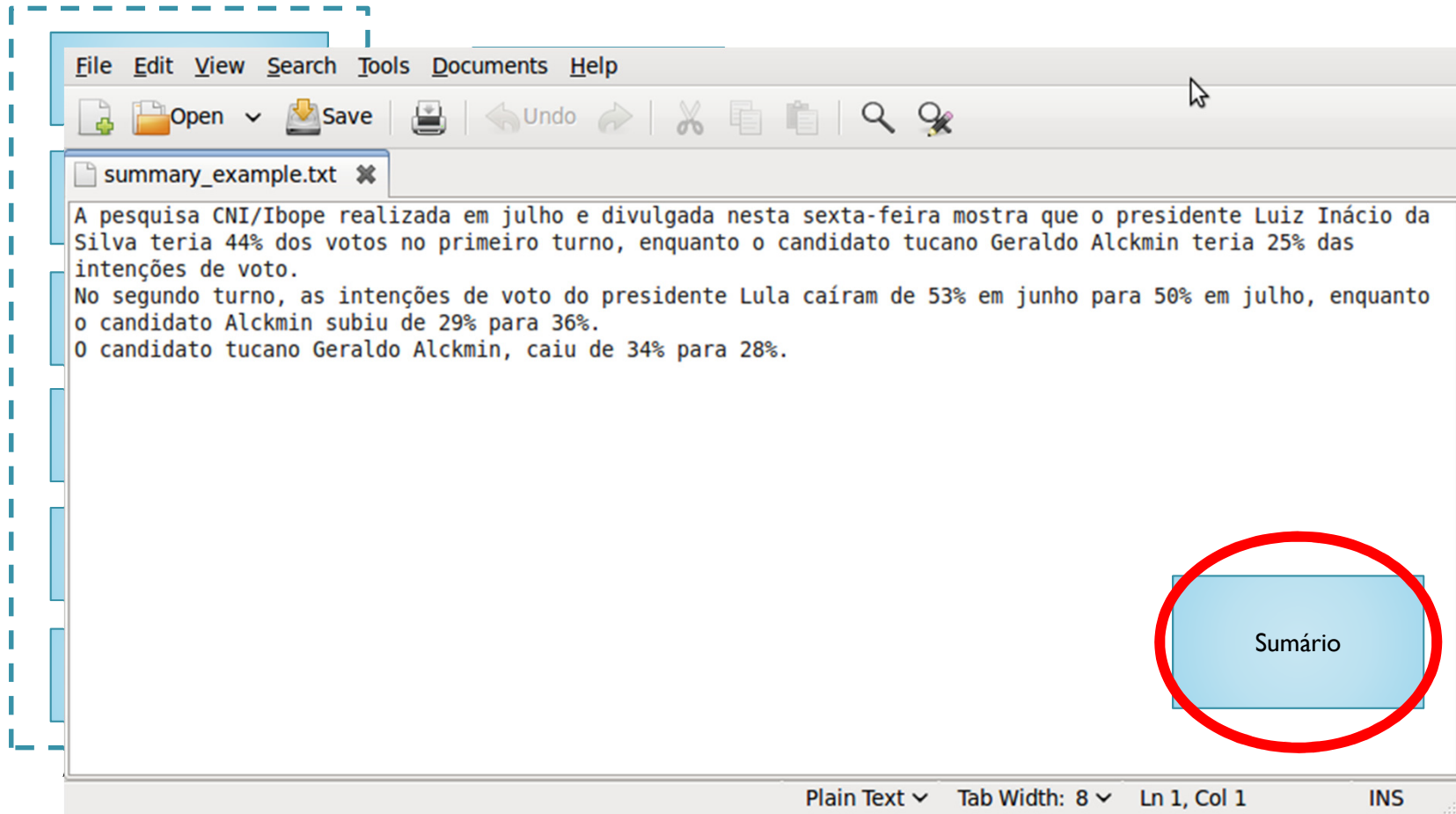
# Arquitetura do Sistema Web

## Síntese



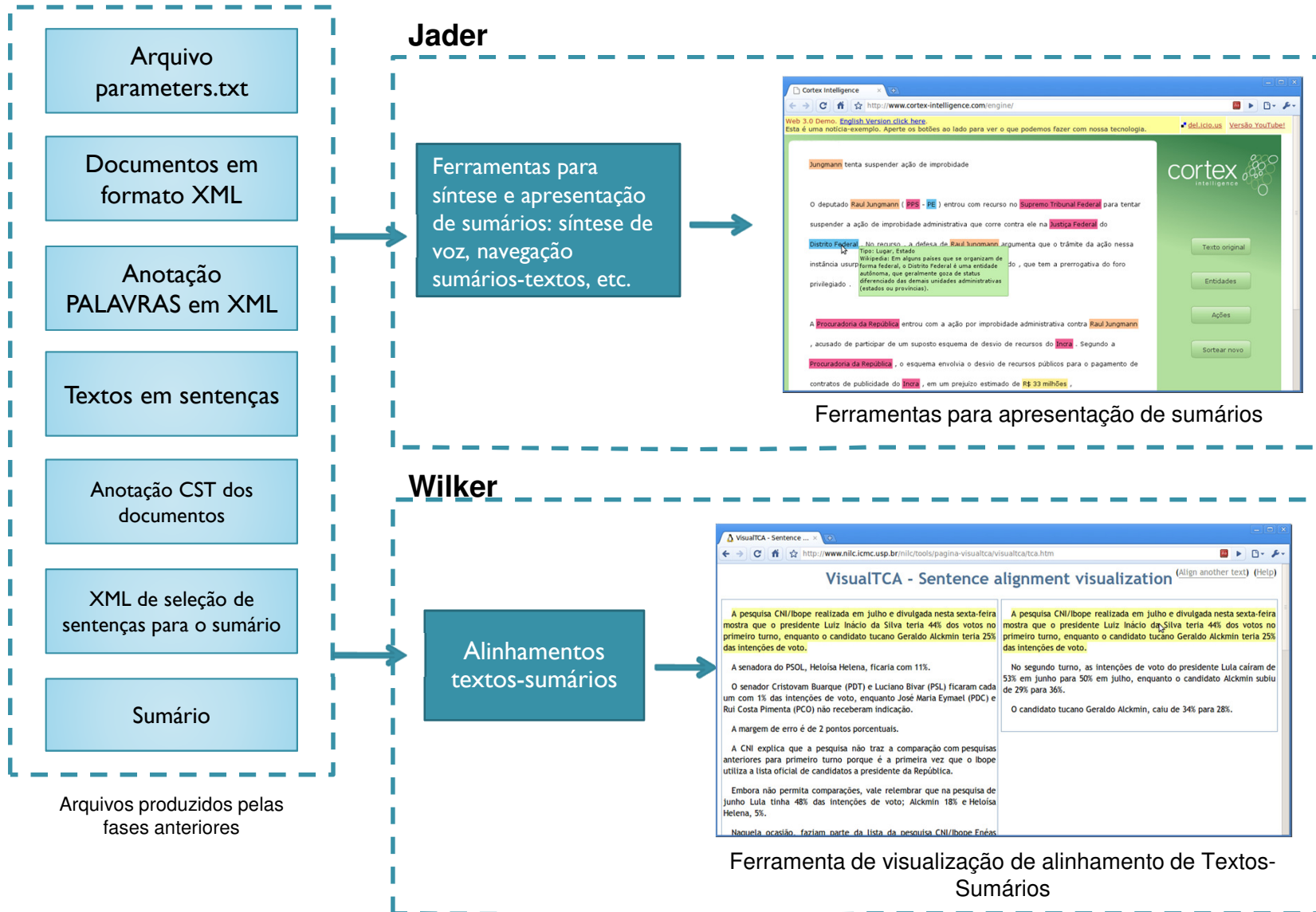
# Arquitetura do Sistema Web


## Síntese



# Arquitetura do Sistema Web

## Apresentação do Sumário Final





**Interface Web** para o projeto: Sumarização  
Automática Multidocumento para o Português do  
Brasil com Base na Teoria de Estruturação  
Multidocumento CST (*Cross-document Structure  
Theory*)